Faceted Classification, Analysis and Search: Some Questions on their Interrelations

Martin Frické

University of Arizona, School of Information, Tucson, AZ 85721, USA,
<mfricke@u.arizona.edu>

Martin Frické is Professor in the School of Information, University of Arizona. He teaches networking, human-computer interaction, logic, and web design, as well as the courses in organization of information, research methods, and information ethics. He typically studies at the interface of information organization, logic, and computer science. At present, he is working on the distributed web and content-based indexing and retrieval (via cryptocurrency infrastructures such as Ethereum smart contracts and the blockchain).

Frické, Martin. 2017. "Faceted Classification, Analysis and Search: Some Questions on their Interrelations." *Knowledge Organization* 44(6): 387-396. 76 references.

Abstract: A description is provided of basic faceted classification, which involves combinations of foci across facets, where the foci within a facet are dependent (i.e., exclusive) and the foci across facets are independent (i.e., orthogonal). This is shown to be suitable for organizing the basic goods that Amazon, the online retailer, sells and for progressive filtering as a mode of search. However, on closer inspection, the Amazon case involves a sorted domain. This is problematic for basic faceted classification. Additionally, books from Amazon would typically carry subject classification, which also is difficult for basic faceted classification. It does not support filtering as a mode of search. Subject classification really requires relatively sophisticated linguistic and logical constructors and modifiers, such as adjectives, adverbs, functions, binary relations, and transitive verbs. These can be part of a synthetic subject classification scheme, but they pose a challenge for faceting.

Received: 25 July 2017; Accepted 27 July 2017

Keywords: facet, concepts, faceted, facets, search, classification, foci

1.0 Introduction

Classification is a matter of concepts. There is no satisfactory account of concepts that is widely or universally accepted. So, without argument or evidence presented here, we will favor an account that makes concepts abstract objects—an account in the tradition of Frege and Tichý (Geach and Black 1980; Tichý 1988). There will be the need to study the relationships between concepts, and this can be done by using first order predicate logic and either simple lambda calculus or the standard set builder notation from set theory (see also Gnoli 2006; Stock 2010). Some examples follow.

Jewelry(x)&Valuable(x)
∃y(Queen(y)&Son(x,y))

are predicate logic formulas; they are actually "open sentences" with free occurrences of the single variable "x."

λx.(Jewelry(x)&Valuable(x))
λx.(∃y(Queen(y)&Son(x,y)))

are lambda abstractions, abstracting using the variable "x."

{x:(Jewelry(x)&Valuable(x))}
{x:(∃y(Queen(y)&Son(x,y)))}

are set builders or set comprehensions, building with the variable "x." If we ignore some subtleties that are not relevant in this context, either the lambda abstractions or the set builders can be understood as denoting concepts—in the examples, denoting the concept of valuable jewelry and the concept of sons of queens.

Plain classification is merely a matter of applying the relevant concept to the entity in question and seeing whether it holds. This is vital, of course, and everyone, from all cultures, does it hundreds of times a day—most times with success. In the context of structure and organization, though, the concepts used are often part of a classification scheme where the concepts used have relations to each other. For example, in the widespread traditional Aristotelian hierarchical taxonomies, where the leaf concepts satisfy the jointly-exclusive-pairwise-distinct (JEPD) property (Frické 2016), the component concepts have to bear certain definite relationships among themselves.

Not every classification is hierarchical, and not every hierarchical classification has the JEPD property (for example, Wikipedia's category "tree," which has cycles, is neither hierarchical nor has the JEPD property). But many classifications do have the JEPD property, and when they do, the items to be classified need to be classified for item retrieval by the maximally specific concepts, i.e., by the "leaves" (or nodes with outdegree zero). Similarly, if classification is envisaged as being paired with search, then there is an important constraint on the way that it is done. Search aims for maximum recall and precision, and it can be proved that the classification act itself on an item needs to be maximally specific to allow this to occur, a result that all catalogers know (Broughton 2004, 115). That is, the "leaves" should be used, i.e., the concepts that have no child concepts. Not every in-the-field classification does this. For example, the *Dewey Decimal Classification* (*DDC*) does not, but it should (Frické 2012, Section 5.4).

Finding out about relationships among concepts can sometimes be a matter of science or mathematics; for example, that horses are animals is for science to determine. But sometimes relationships can be a matter of logic and reasoning. For example, the concepts of man and young man are related. The lambda abstractions below indicate that "young man" is a sub-concept of the concept "man."

λx.(Man(x))
λx.(Man(x)&Young(x))

In this case, a hierarchical relationship can be established by proving that being a man follows from being a young man. These relationships exploit the fact that some concepts are compound concepts, i.e., they have components. Entire classification schemes or schedules can be produced using this. That is exactly what logical division does (Frické 2016).

Faceted classification typically uses compound concepts, which are seen as being composed from atomic concepts. It rests on the fact that concepts can have categories or kinds (Vickery 1960, 1966; Willetts 1975; Austin 1984; Foskett 1996; Morville and Rosenfeld 2006; Lambe 2007; Cheti and Paradisi 2008; Slavic 2008; La Barre 2010b). For example, the "being female concept" is an instance of "being a gender classifying concept," and the "being young concept" is an instance of "being an age classifying concept." The "kinds" of concepts are "facets," and the "terminal or maximally specific atomic concepts" within a facet are "foci" (Ranganathan 1959, [1937] 1967; Vickery 1960, 1966, 2008; Gardin 1965; Buchanan 1979; Foskett 1996, 2003; Broughton 2004, 2006; La Barre 2006, 2010a; Wilson 2006; Gnoli 2008). One assumption that can be made is that the maximally specific concepts within the same facet are "exclusive" (that, for example, if a person is young then they are not, at one and the same time, old). Sometimes it is also helpful if the maximally specific concepts are "exhaustive"—that is, always at least one of them applies. As mentioned, the classification act itself on an item needs to be maximally specific, i.e., it is the foci that should be used for item classification. There can be more general concepts within a facet schedule, but these would be used for organization or for directing a search and not for the actual labeling of an item. There is another technical feature that is useful to have. It is that the facets be "orthogonal." This means that a choice of a focus from one facet is entirely independent of a choice of focus from another facet (e.g., a man can be either young, middle-aged, or old, i.e., that something is a man does not *ipso facto* mean, say, that he is old). Foci are "dependent" within a facet (thanks to exclusivity) and "independent" across facets (thanks to orthogonality). The resulting classificatory structure is that of a directed acyclic graph (DAG). This is different to a classificatory tree or plain hierarchy, because it is possible for a concept (or node) to have two parents. The concept "λx.(Female(x)&Young(x))," for example, has two parents—it is a subconcept (a "child") of both "λx.(Female(x))" and of "λx.(Young(x))." The fact that the structure is a DAG (and is finite) means that there are nodes that do not have children. These are the narrowest most specific concepts. Effectively these are "leaves" of the DAG and it is these that are used for the actual classifying. Each person, in a miniature example scheme, might be classed as exactly one of

λx.(Female(x)&Young(x))
λx.(Female(x)&Middleaged(x))
λx.(Female(x)&Old(x))
λx.(Male(x)&Young(x))
λx.(Male(x)&Middleaged(x))
λx.(Male(x)&Old(x)).

It helps for clarity of exposition to imagine that the classified people are labeled or tagged with a tag depicting their classification. Each will carry exactly one of these six tags.

Modern computing and algorithms, when addressing data, has pared down to using what might be called universal data structures for data (see, for example, JSON 2017). One of these is the key-value pair, and, in particular, collections, without order, of key-value pairs, and these are sufficient to characterize items or objects. With basic faceted classification, the keys will be the facets (e.g., gender, age), and the values will be the foci (e.g., male, old). So, a particular classification for an individual item might look like

> {Gender: male, Age: old}.

That the facets are orthogonal means additional facets can be added without disturbing any classification or schema that is already in place, for example

> {Gender: male, Age: old, PlaceOfResidence: London, …}.

Additionally, for flexibility, facets (or particular key-value pairs) can be omitted for some items classified under a scheme. Parallel to this, any collection of key-value pairs could be entered as rows and columns in a relational database (perhaps with the assistance of a "null" value for undefined values).

A faceted classification would ordinarily be polyhierarchical. The faceting allows a vast number of combinations of foci across the facets, which is an advantage. Were an attempt made to put these combinations into a non-faceted single hierarchy, there would be questions of which facet goes higher up the hierarchy (e.g., gender or placeofresidence?), and there would be much repetition of structure because of "distributed relatives" (Savage 1946). For instance, "male" would have to have below it, say, "London," "Bristol," etc., but then so too would "female,"

Nowadays, with computers, faceted classification goes hand in hand with faceted search (Tunkelang 2009). Each search aims to produce a descriptor, which is essentially a combination of concepts—for example, "λx.(Female(x))," or, "λx.(Male(x) or Old(x))." Then the search finds the items that are tagged satisfying the descriptor. Search can be an iterative process. For example, a user can search for males, then search among the males for the old males. This is done by narrowing using concepts—for example, λx.(Male(x)). These are not usually themselves tagging combinations of foci, but they can be used to select foci. These kinds of searches can be Boolean searches such as searching, for example, for "man AND old," "woman AND NOT old," etc., and the Boolean searches can perhaps be aided by computer interfaces that present the facets and their foci. Process searches are almost always "filtering," that is they narrow on the results to hand. The following is a typical description of the process (Arenas et al. 2014):

> Faceted search is a prominent approach for querying document collections where users can narrow down the search results by progressively applying filters, called *facets*. A facet typically consists of a property (e.g., 'gender' or 'occupation' when querying documents about people) and a set of possible string values (e.g., 'female' or 'research'), and documents in the collection are annotated with property-value pairs. During faceted search users iteratively select facet values and the documents annotated according to the selection are returned as the search result.

Also, in terms of search, the individual facets can have, and usually would have, a structure. So, for example, the first order predicate "λx.(HavingAGender(x))" could be introduced to provide a root for the gender facet, and then the facet would have a hierarchical structure with male and female being sub-concepts of the root. The individual facets themselves might be organized in different ways. As Barbara Kwasnik writes (1999, 39-40),

> Each facet can be developed/expanded using its own logic and warrant and its own classificatory structure. For example, the Period facet can be developed as a timeline; the Materials facet can be a hierarchy; the Place facet a part/whole tree, and so on.

There is a point that arises here for advanced treatments. If the type of classificatory structure is different across facets, there might not be a single ordering relation in use for the compound concepts. For example, the relationship of sub-concept is being used in the present treatment, but if a place facet was a part/whole tree, something different would have to be invoked, because a part is not, strictly speaking, a sub-concept of a whole.

This basic faceted classification and faceted search has been called "ersatz" faceting (Frické 2013). Ersatz faceting has the following features:

- there is one domain,
- selection by a single focus, or combination of foci from different facets, merely identifies a subset of the domain,
- selection of a single focus during a search automatically rules out any choice of other foci from the same facet, by exclusivity, so the other foci no longer need to be offered as choices in a search interface of the ongoing search,
- the order in which narrowing or filtering operations are carried out is inconsequential as to the final resulting subset (i.e., the operations permute or are symmetric),
- that the operations permute means that the syntax for a faceted description language is relatively open (e.g., either of "old men" or "men old" would be fine).

- the narrowing operations lend themselves to representation as Boolean search (in particular ANDing the operations in any order),
- for Boolean search
  - the ANDing operations need only AND foci across different facets (because choice of a focus within a facet implicitly and automatically ANDs that choice with NOT of all the other foci within a facet),
  - OR operations across facets is semantically sound (for example, "men OR old" identifies a subset of people),
  - OR operations within a facet is semantically sound (for example, "young OR old" identifies a subset of people),
  - in fact, all the Boolean operations are sound, provided that the mutual exclusivity of foci within a facet is respected.

There are formal accounts of faceted classification, using set theory, formal concept analysis, lightweight ontologies, and (mathematical) category theory, etc. (Priss 2000; Giunchiglia et al. 2007; Wei et al. 2013; Harris 2016). All of these use one-place properties or attributes, exclusive foci and orthogonal facets, i.e., they all model ersatz faceting.

From a logical point of view, ersatz faceting ANDs the conditions or attributes. It uses what Ranganathan (1960) called simple "superimposed" classes (or topics). Other labels for the compound concepts might be "intersective" types (because the compounds generate set intersections) or "conjunctive" types (because the logic uses AND to form the compounds).

2.0 An example: classifying the goods and services of the online retailer Amazon

2.1. Goods

Amazon acquires, organizes, stores, advertises, eases access to, and sells many things. How might this be done? How might the user be helped in finding, say, MacBook Airs? One obvious possibility is to use ersatz faceting, faceted search, Boolean search, free search, natural language search, and thesaurus support. There could be a single facet of, perhaps, all of Amazon's goods with a standard hierarchy of those, coming down to electronic goods, coming down to computers, laptop computers, etc. to MacBook Airs. Then a second facet might be price with a hierarchy coming down to expensive. So, the MacBook Air type would itself carry the single classification tag of $\lambda x.(Laptop(x)\&Expensive(x))$. Maybe something more elaborate would be better, in terms of more facets. There might be an audience facet, a functional purpose facet, a gift occasion facet, and so on. But, all the while, the Air type effectively would carry a single tag.

Search could be directly via a search box, or via facets, which would provide filtering. There would need to be thesaurus support (because, for example, some users might describe "laptops" as being "portables"). A user could also speak to Alexa to get natural language help.

In sum, ersatz faceting seems adequate for websites, and similar, which give the user selection by progressive filtering via orthogonal properties (c.f., for example, Endeca, Flamenco, and Apache Solr (Hearst 2008; Zelevinsky et al. 2008; Tunkelang 2009; Smiley and Pugh 2011; Wei et al. 2013).

2.2. Services

There is a problem, though. Amazon does not sell just goods; it also sells services. Users of the web may not realize this, but a large number of websites are run, in whole or part, by Amazon, e.g., AirBNB, Netflix, Zillow (Amazon 2017). A large company like AirBNB, say, might not feel the need to actually own computers, web servers, disk drives, etc., so they instead go to Amazon looking for devices "aaS" ("as a Service"). So, they might buy a computer as a service.

At this point, we need to sit up and pay attention. A computer as a service is not a computer (and it will not have a physical location or a barcode). Ersatz faceting is not going to work. A search cannot start with Amazon's physical goods and filter down to an item which is not a physical good. With ersatz faceting, any predicate or conceptual component in the faceted scheme can, potentially at least, apply to any of the items being classified. But this no longer holds. Take, for example, the concept "available for pickup in your zip code" which applies to goods but not to services.

Effectively there is a sorted domain, a domain of goods, a domain of services, and maybe other domains (Blasius et al. 1989; Sowa 2000a, b). How many sorts might there be in the domain? This is hard to say and hard to find out. One approach would be to take a general ontology, say PROTON (PROTo ONtology), or one of the other general ontologies (Khoo and Na 2006; Mascardi et al. 2007), and let that tell us. PROTON is a lightweight upper-level ontology, serving as a modeling basis for a number of tasks in different domains (see, http://ontotext.com/products/proton/). PROTON has about 300 classes, and about 25 higher level sorts—we might expect that Amazon's domain would have a rich partition into sorts. Of course, Amazon will not sell even most of the kinds of things that these ontologies invoke, but it might sell many.

There could be separate faceted schemes for the separate domains. But it would be challenging to produce one single scheme for all. The question to be asked is "do we want the different sorts to be in the same facet, or in different facets?" Putting them in the same facet is superficially attractive in that the exclusivity leads to the choice of one sort when classifying, but it is awkward in combination with other facets. For example, a person from the sort facet might combine with the focus "unmarried" but can an Amazon service from the sort facet be "unmarried?" Sorts arise in the first place, because the sorts are not the "same." So, why would they combine with the same foci from the same other facets? Putting them in different facets is also problematic, because the orthogonality requires that they combine freely with each other, but sorts, so-to-speak, are intrinsically different.

Also, roughly speaking, ontologies tell you primarily what "kinds" of things there are, "kinds" here meaning "kinds essentially," but faceted classification and search is equally interested in "kinds accidentally." For example, being a computer might be an essential kind of an item, whereas being large or small or expensive or inexpensive would be accidental features that the computer might have. But higher level ontologies would not tell us about these.

In sum, prospects do not look brilliant for a single faceted classification of "all" the goods, services, etc. that Amazon sells. The core problem is with getting the facets to be orthogonal and the foci within the individual facets to be exclusive and exhaustive. There could be miniature localized faceted sub-classifications within a wider schema. But, these would be piecemeal and eclectic, and there might be much reduplication (e.g., "distributed relatives"). If one key in the key-value collection is for sort, then essentially there are separate faceted classifications for each sort and not one classification for all.

2.3. Books

Amazon, as is well known, also sells books. What is different about books is that they can be classified by topic or subject, i.e., given subject classification. Amazon books will carry a number of metadata fields, e.g., publisher, price, etc. and those metadata fields might be conceived of by some as being "facets" (they certainly are key-value pairs). But, the "values" of the subject field itself might themselves be faceted thus forming distal facetings (Clarkson et al. 2009). Some standard subject headings systems have many entries. For example, the *Library of Congress Subject Headings* (*LCSH*) has over 300,000 headings. It would be preferable not to have 300,000 foci as separate entries on their own. It would be better to construct them as compounds out of other facets with a small number of atomic foci, and, indeed, there have been such attempts. There is also the semantic annotation of documents by computer scientists, which is similar to the librarian subject classification (Deyab et al. 2016).

The first move here with subject headings, or topics, or semantic annotations, is to use the "referential semantics" of subject languages (Svenonius 2000, 130). The concepts, e.g., old male, are not now being used to classify people, rather the concepts are tags to identify topics. So, old males might be a topic of a book. And books can have multiple topics, so they might carry multiple tags. The multiple tags can be "inconsistent," unlike the case with physical goods. For instance, a book can be on the topics "old men" and "young women." In addition, perhaps a single tag should be able to be "inconsistent." For example, the following possible tag seems perfectly good, but notice that it has in it a combination of two foci from the same facet

#young and old women.

Then topics range wider than mere concepts as classifying devices. Topics can be plain entities (e.g., named people or places, like London), and topics can be statements, true or false, (e.g., "Brexit will have severe economic consequences").

Topics typically concern "aspects." Brian Vickery writes (1975, 9)

> Taxonomy is basically concerned with classifying "natural kinds"—of organisms, of soils, of substances. Documentation has to classify what is written about these objects, and must take into account not only the natural kinds but also their properties, behaviour, interactions, and operations on them.

Explanations of "aspects" can be found in Mills and Broughton (1977), Broughton (2004), Broughton et al. (2005) and Hjørland (2006).

Topics, as aspects, are going to be compound concepts, usually compound nouns. However, from a logical point of view, they are typically going to involve more than conjunction in their construction. Consider

the habitat of rabbits.

This is an "aspect" of rabbits, but its logical form is not

λx.(Habitat(x)&Rabbits(x)) /* 'habitats which are also rabbits' */.

Rather, it needs to be expressed using a function

λx.(x=habitatOf(rabbits)).

This is not now ersatz faceting—the collective habitats of rabbits is not a narrowing or filtering of rabbits.

Subjects are often more complex yet in form (Vickery 2008, 147-8):

> For example:
> Ecological study of the skin colour of desert animals
> Streptomycin therapy for osteomyelitis of femur
> Effect of humus on crumb formation in loamy sand
> Radiographic diagnosis of bone cancer
> Grey wool ankle socks for hiking
> Damages for personal injury in English law
> Plating the spokes of bicycle wheels
> Prevention of mould spoilage of parboiled rice in storage in silo
> Wind tunnel measurement of Reynolds number for boundary layer transition in model aircraft

Many of these are not "aspects"—they are simply complex concepts which require a range of features from logic. Such compound concepts can certainly be rendered in first order logic. But there is a question of whether the concepts' components or atoms can be divided into facets, and there is the telling observation that conjunction or composition will be inadequate. In ersatz faceting, the foci are conjoined and it is this that allows the user to filter, narrow, and refine by, for example, going from men to young men. But consider a simple compound concept like "sons of queens." This is not constructed by plain conjunction. Assume it could be made somehow from a facet for royalty and a separate facet for "relations of." A search for books on the focus topic of queens would produce a subset of books, but then adding, a relation focus of "son of" to that topic, and result set likely would not filter on the initial results. Simply put, documents on sons of queens is not a subset of documents on queens.

To sum up, examples such as Amazon teach us that simple faceting is not going to work in advanced cases.

3.0 Subject classification, facet analysis and the construction of faceted classifications

Facet analysis is the process of producing a faceted classification for a discipline or area of study. The initial theory (or theories) for facet analysis, comes from Ranganathan and the United Kingdom Classification Research Group (CRG) (Ranganathan [1937] 1967; Classification Research Group 1955; Gopinath 1992; Spiteri 1998; Broughton 2004). Some problems these originators studied can be set aside. For example, to produce an overall linearization of the subjects to shelve books, they looked at the order of the facets in a compound concept, so-called "citation order," and the order of foci within the facets, known as "filing order." But shelving books is no longer central. There are a number of relatively modern accounts of facet analysis discussed by, for example, Soergel (1974), Mills and Broughton (1977), Buchanan (1979), Kwasnik (1992, 1999), Aitchison et al. (2000), Prieto-Díaz (2003), Broughton (2004), La Barre (2004), Mills (2004), Gnoli (2008), Vickery (2008) and Hjørland (2012). Computer scientists have also been very active in this area (Wei et al. 2013).

Let us provide a quick and incomplete description of facet analysis. It usually works with written or spoken pieces of language, not concepts. It addresses terms, i.e., nouns, nominative phrases, etc. Selective domain analysis is done on titles, contents, and communities, because the classification is domain dependent. The terms are clustered. Prominent clusters become facets. Exemplars from the facets are chosen as foci. Then, faceted classification is done using these. For example, according to Vickery (2008, 156), the topic

> the fortification with vitamins of infant foods made from cereal flour

might be classified

> Technology: food
> Product: cereal
> Product for: infant
> Product form: flour
> Product processing: fortification by vitamin.

Then a hierarchy or structure can be imposed on the facets and thesaurus support added to allow for differences in vocabulary.

More generally, a relatively complete list of CRG general facet categories (i.e., proto-facets), which would be shaped to meet the requirements of the individual disciplines, is

> Thing
> Kind
> Part (organ, constituent)
> Property
> Material
> Process (an action internal to the item)
> Operation (an action performed on the item)
> Patient (object of action, raw material)
> Product (substance)
> By-product
> Agent
> Space
> Time.

As mentioned earlier, computer scientists would typically compose the foci by simple conjunction. For example, see the term composition algebra of Tzitzikas et al. (2004). This leads to ersatz faceting. But notice that a typical library science facet analysis would permit non-conjunctive types. For example, it might combine the kind "human" with the process "growth" to arrive at the topic "human growth"—the kinds come from predicates and the processes from functions.

Researchers in librarianship (Farradane 1950, 1963; Moss 1964; Gardin 1965; Perreault 1965; Gardin 1969; Perreault 1969; Gardin 1973; Austin 1984) have been very sophisticated in what they have devised. They have faceted down to foci but allow simple relations, which can be functions, between foci. So that, for example, "explosions" might be one focus, "injuries" another, and "…causing…" a relation between certain kinds of foci, and these components could be put together to form the subject tag "explosions causing injuries."

There is a richness here that is not present in ersatz faceting. There are functions, relations and basically the full range of logical connectives, but this comes at a price as far as faceting is concerned. As an illustration, the relations themselves between foci need organizing. Here are some relations:

… causing …
… correlating with …
… preventing …
… giving …
… donating …
… loving …

or, more generally, for example, constructions expressed by transitive verbs or comparative adjectives. How are these to be organized or faceted? It is an open question.

Where this research also is a little lacking, if it may be phrased that way, is that by happenstance it largely predates powerful computers, the Internet, and faceted search. It seeks to build pre-coordinated controlled subject headings by synthesis from faceted components. There might be, or mighty have been, post-coordinated faceted systems lurking in the background, perhaps inspired by Mortimer Taube's uniterm work (Taube 1951, 69), but the actual subject headings as final products were pre-coordinated. But, now, in 2017, the user may be more inclined to employ faceted search as an interactive, sequential, and filtering operation which is post-coordinated. The input is focus by focus, usually one at a time, and the computer has to make sense of it. Right now, the sophisticated faceted classification of librarians and the faceted search of computer scientists do not match very well with each other.

4.0 A positive suggestion

Here is a positive suggestion. The main problem with simple faceting is that it uses conjunction (AND) as its main constructor for combining facets and foci, but there is a better and more powerful way. It is that of using functions as the atomic components, and function applications and function composition as the constructors. This is what would be done in formal linguistics (Jacobson 2014). It is what would be done in a Fregean approach to concept analysis (Geach and Black 1980; Tichý 1988). And it is what would be done in functional programming in computer science such as that of Haskell.org (https://www.haskell.org/). Haskell has a type system and mechanisms for combining functions and types. A propos this positive suggestion, one should acknowledge that the making of a suggestion and carrying it through are only distantly related.

5.0 Conclusion

Computer scientists enamoured with faceted search do not pay enough attention to the work of librarians. In particular, they tend to restrict themselves to ersatz faceting (i.e., Ranganathan's superimposed classes), and this is not enough in the general case. Librarians designing faceted subject heading languages do not pay enough attention to faceted search. In particular, little attention is paid to how such searches might be interactive and illuminating, yet likely without having the property of filtering. A good example of close-to-being-faceted subject headings are the medical subject headings of *MeSH*(www.nlm.nih.gov/mesh/). *MeSH* is wonderful, one of the best subject heading languages and systems, but its support for search is baroque (Lowe 1994). The logical structure of annotating topics is relatively involved. There are increasingly impressive natural language parsers that can determine linguistic and logical structure. Librarians, working with computers as assistants have an opportunity here. The research needs to explore functions, relations, and more advanced logical structures. The result may be improved subject heading languages (but ones that might not be faceted in their entirety). Synthesis should be possible, but faceting is a more demanding target.

References

Aitchison, Jean, Alan Gilchrist, and David Bawden. 2000. *Thesaurus Construction and Use: A Practical Manual.* 4th ed. Chicago: Fitzroy Dearborn.

Amazon. 2017. "AWS Customer Success." https://aws.amazon.com/solutions/case-studies/

Arenas, Marcelo, Bernardo Cuenca Grau, Evgeny Kharlamov, Sarunas Marciuska, and Dmitriy Zheleznyakov. 2014. "Faceted Search over Ontology-Enhanced RDF Data." In *CIKM'14: Proceedings of the 2014 ACM International Conference on Information and Knowledge Management, November 3-7, 2014, Shanghai, China*. New York: ACM, 939-48. doi:10.1145/2661829.2662027

Austin, Derek. 1984. *PRECIS: A Manual of Concept Analysis and Subject Indexing*, 2nd ed. London: The British Library.

Blasius, K. H., U. Hedstuck, and C.R. Rollinger, eds. 1989. *Sorts and Types in Artificial Intelligence*. Lecture Notes in Computer Science 418. Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag.

Broughton, Vanda. 2004. *Essential Classification*. New York: Neal-Schuman.

Broughton, Vanda. 2006. "The need for a faceted classification as the basis of all methods of information retrieval." *Aslib Proceedings* 58: 49-72.

Broughton, Vanda, Joachim Hansson, Birger Hjørland, and M. J. López-Huertas. 2005. "Knowledge Organization." In *European Curriculum Reflections on Library and Information Science Education*, ed. Leif Kajberg and Leif Lørring. Copenhagen: Royal School of Library and Information Science, 133-48.

Buchanan, Brian. 1979. *Theory of Library Classification*. Outlines of Modern Librarianship. London: Clive Bingley.

Cheti, Alberto, and Federica Paradisi. 2008. "Facet Analysis in the Development of a General Controlled Vocabulary." Axiomathes 18: 223-41.

Clarkson, Edward C., Shamkant B. Navathe, and James D. Foley. 2009. "Generalized Formal Models for Faceted User Interfaces." In *JCDL '09: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries, Austin, TX, USA, June 15-19, 2009*. New York: ACM, 125-34. doi:10.1145/1555400.1555422

Classification Research Group. 1955. "The Need for a Faceted Classification as the Basis of all Methods for Information Retrieval." *Library Association Record* 57, no. 7: 262-8.

Deyab, Mohamed Fathy, Hisham Emal El-Din ElShishiny, Ahmed Ragheb, and Ahmed Mousaad Abdel Wanees. 2016. Intelligent Text Annotation. US Patent 9292483 B2, filed June 25, 2008, and issued March 22, 2016.

Farradane, J. E. L. 1950. "A Scientific Theory of Classification and Indexing." *Journal of Documentation* 6:83-99.

Farradane, J. E. L. 1963. "Relational Indexing and New Methods of Concept Organization for Information Retrieval." In *American Documentation Institute 26th Annual Meeting*. Washington: American Documentation Institute, 135-6.

Foskett, A. C. 1996. *Subject approach to information*. 5th ed. London: Library Association Publishing.

Foskett, Douglas J. 2003. "Facet Analysis." In *Encyclopedia of Library and Information Science*, ed. Miriam A. Drake. New York: Marcel Dekker, 1063-7.

Frické, Martin. 2016. "Logical Division." *Knowledge Organization* 43: 539-49.

Frické, Martin. 2012. *Logic and the Organization of Information*. New York: Springer.

Frické, Martin H. 2013. "Facets: Ersatz, Resource and Tag." *Information Research: An International Electronic Journal* 18, no. 3.

Gardin, Jean-Claude. 1965. "Free Classifications and Faceted Classifications: Their Exploitation with Computers." In *Classification Research: Proceedings of the Second International Study Conference held at Hotel Prins Hamlet, Elsinore, Denmark, 14th to 18th September 1964*, ed. Pauline Atherton. Copenhagen: Munksgaard, 161-76.

Gardin, Jean-Claude. 1969. "Semantic Analysis Procedures in the Sciences of Man." *Social Science Information* 8, no. 1: 17-42.

Gardin, Jean-Claude. 1973. "Document Analysis and Linguistic Theory." *Journal of Documentation* 29: 137-68.

Geach, Peter, and Max Black. 1980. *Translations from the Philosophical Writings of Gottlob Frege*, ed. Peter Geach and Max Black. 3rd ed. Oxford: Blackwell.

Giunchiglia, F., M. Marchese, and I. Zaihrayeu. 2007. "Encoding Classifications into Lightweight Ontologies." *Journal on Data Semantics* 8: 57–81.

Gnoli, Claudio. 2006. "The Meaning of Facets in Nondisciplinary Classifications." In *Knowledge Organization for a Global Learning Society: Proceedings of the Ninth International ISKO Conference 4-7 July 2006 Vienna, Austria,* ed. Gerhard Budin, Christian Swertz, Konstantin Mitgutsch. Advances in Knowledge Organization 10. Würzburg: Ergon Verlag, 11-8.

Gnoli, Claudio. 2008. "Facets: A Fruitful Notion in Many Domains." *Axiomathes* 18: 127-30.

Gopinath, M.A. 1992. "Ranganathan's Theory of Facet Analysis and Knowledge Representation." *DESIDOC Journal of Library and Information Technology* 12, no. 5: 16-20.

Harris, Daniel R. 2016. "Foundations of Reusable and Interoperable Facet Models using Category Theory." *Information Systems Frontiers* 18: 953-65.

Hearst, Marti A. 2008. "UIs for Faceted Navigation: Recent Advances and Remaining Open Problems." In *HCIR 2008: Proceedings of the Second Workshop on Human-Computer Interaction and Information Retrieval Microsoft Research, Redmond, Washington, USA October 23, 2008*, 13-17. http://ryenwhite.com/hcir2008/doc/HCIR08-Proceedings.pdf

Hjørland, Birger. 2006. "Aspect or Discipline versus Entity or Phenomena or 'One Place' Classification." Unpublished manuscript, formerly in: http://www.iva.dk/bh/lifeboat_ko/concepts/aspect_classification.htm

Hjørland, Birger. 2012. "Facet Analysis: The Logical Approach to Knowledge Organization." *Information Processing and Management* 49: 545-57. doi: 10.1016/j.ipm.2012.10.001

Jacobson, Pauline. 2014. *Compositional Semantics: Introduction to the Syntax/Semantics Interface*. Oxford Textbooks in Linguistics. New York: Oxford University Press.

JSON. 2017. Javascript Object Notation. Accessed March 20. http://www.json.org/

Khoo, Christopher S. G., and Jin-Cheon Na. 2006. "Semantic Relations in Information Science." *Annual Review of Information Science and Technology* 40: 157-228. doi:10.1002/aris.1440400112

Kwasnik, Barbara H. 1992. "The Legacy of Facet Analysis." In *S.R. Ranganathan and the West*, ed. Ravindra Nath Sharma. New Delhi: Sterling, 98-111.

Kwasnik, Barbara H. 1999. "The Role of Classification in Knowledge Representation and Discovery." *Library Trends* 48: 22-47.

La Barre, Kathryn. 2004. "Adventures in Faceted Classification: A Brave New World or a World of Confusion?" In *Knowledge Organization and the Global Information Society: Proceedings of the Eighth International ISKO Conference, 13-16 July 2004 London, UK*, ed. Ia C. McIlwaine. Advances in Knowledge Organization 9. Würzburg: Ergon Verlag, 79-84.

La Barre, Kathryn. 2006. "The Use of Faceted Analytico-Synthetic Theory as Revealed in the Practice of Website Construction and Design." PhD diss., Indiana University.

La Barre, Kathryn. 2010a. "Facet Analysis." In *Annual Review of Information Science and Technology* 44: 243-86. doi:10.1002/aris.2010.1440440113

La Barre, Kathryn. 2010b. "A Semantic (Faceted) Web?" *Les cahiers du numérique* 6, no. 3: 103-31.

Lambe, Patrick. 2007. *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness*. Chandos Knowledge Management Series. Oxford: Chandos Publishing.

Lowe, Henry J., and G. Octo Barnett. 1994. "Understanding and Using the Medical Subject Headings (MeSH) Vocabulary to Perform Literature Searches." *Journal of the American Medical Association* 271: 1103-8.

Mascardi, Viviana, Valentina Cordì, and Paolo Rosso. 2007. "A Comparison of Upper Ontologies." In *WOA 2007: dagli oggetti agli agenti, agenti e industria, applicazioni tecnologiche degli agenti software*, ed. Matteo Baldoni, Antonio Boccalatte, Flavio De Paoli, Maurizio Martelli and Viviana Mascardi. Turin: Seneca Edizioni 2007, 55-64.

Mills, Jack. 2004. "Classification and Logical Division in Information Retrieval." *Library Trends* 52: 541-70.

Mills, J., and Vanda Broughton. 1977. *Introduction and Auxiliary Schedules*. Vol. 1 of Bliss *Bibliographic Classification*. 2nd ed. London: Butterworths.

Morville, Peter, and Louis Rosenfeld. 2006. *Information Architecture for the World Wide Web*, 3rd ed. Sebastapol, CA: O'Reilly.

Moss, R. 1964. "Categories and Relations: Origins of Two Classification Theories." *American Documentation* 15: 296-301.

Perreault, J. M. 1965. "Categories and Relators: A New Schema." *Revue internationale de documentation*, 32: 136-44.

Perreault, J. M. 1969. *Towards a theory for UDC*. Hamden, CT: Archon Books & Clive Bingley.

Prieto-Díaz, Rubén. 2003. "A Faceted Approach to Building Ontologies." In *Information Reuse and Integration 2003 (IRI 2003): Proceedings Fifth IEEE Workshop on Mobile Computing Systems and Applications*. USA: IEEE, 458-65.

Priss, Uta. 2000. "Faceted Information Representation." In *Working with Conceptual Structures: Proceedings of the 8th International Conference on Conceptual Structures; Logical, Linguistic, and Computational Issues, Darmstadt, August 14-18, 2000*, ed. Gerd Stumme. Aachen: Shaker Verlag, 84-94.

Ranganathan, S. R. (1937) 1967. *Prolegomena to Library Classification*. 3rd ed. Ranganathan series in library science 20. London: Asia Publishing House.

Ranganathan, S. R. 1959. *Elements of Library Classification*, 2nd ed. London: Association of Assistant Librarians.

Ranganathan, S. R. 1960. *Colon Classification: Basic Classification*, 6th ed. Ranganathan Series in Library Science 4. London: Asia Publishing House.

Savage, Ernest. 1946. *Manual of Book Classification and Display*. Library Association Series of Library Manuals 8. London: Allen & Unwin.

Slavic, Aida. 2008. "Faceted Classification: Management and Use." *Axiomathes* 18: 257-71.

Smiley, David, and Eric Pugh. 2011. *Apache Solr 3 Enterprise Search Server*, 2nd ed. Birmingham: Packt Publishing.

Soergel, Dagobert. 1974. *Indexing Languages and Thesauri: Construction and Maintenance*. Los Angeles: Melville.

Sowa, John F. 2000a. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks/Cole.

Sowa, John F. 2000b. "Ontology, Metadata, and Semiotics." In *Conceptual Structures: Logical, Linguistic, and Computational Issues: 8th International Conference on Conceptual Structures, ICCS 2000, Darmstadt, Germany, August 14-18, 2000*, ed. B. Ganter and G. W. Mineau, Lecture Notes in Computer Science 1867. Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin, 55-81.

Spiteri, L. 1998. "A Simplified Model for Facet Analysis: Ranganathan 101." *Canadian Journal for Library and Information Science* 23: 1-30.

Stock, Wolfgang G. 2010. "Concepts and Semantic Relations in Information Science." *Journal of the American Society for Information Science and Technology* 61: 1951–69.

Svenonius, Elaine. 2000. *The Intellectual Foundation of Information Organization*. Digital Libraries and Electronic Publishing. Cambridge, MA: M.I.T. Press.

Taube, Mortimer. 1951. "Functional Approach to Bibliographic Organization: A Critique and a Proposal." In *Bibliographic Organization: Papers Presented Before the Fifteenth Annual Conference of the Graduate Library School, 24-29 July 1950*, ed. Jesse H. Shera and Margaret Egan. Chicago: University of Chicago Press, 57-71.

Tichý, Pavel. 1988. *The Foundations of Frege's Logic*. Foundation of Communication. Berlin: Walter de Gruyter.

Tunkelang, Daniel. 2009. *Faceted Search*. Synthesis lectures on information concepts, retrieval, and services 5. San Rafael: Morgan and Claypool, 1-80.

Tzitzikas, Yannis, Anastasia Analyti, Nicolas Spyratos, and Panos Constantopoulos. 2004. In *Information Modelling and Knowledge Bases XV*, ed. Yasushi Kiyoki. Frontiers in Artificial Intelligence and Applications 105. Amsterdam: IOS Press, 67-87.

Vickery, Brian Campbell. 1960. *Faceted Classification: A Guide to Construction and Use of Special Schemes*. London: Aslib.

Vickery, Brian Campbell. 1966. "Faceted Classification Schemes." In *Rutgers Series on Systems for the Intellectual Organization of Information*, ed. S. Artandi. Rutgers Series on Systems for the Intellectual Organization of Information 5. New Brunswick, NJ: Graduate School of Library Service, Rutgers, The State University.

Vickery, Brian Campbell. 1975. *Classification and Indexing in Science*. 3rd ed. London: Butterworths.

Vickery, Brian Campbell. 2008. "Faceted Classification for the Web." *Axiomathes* 18: 145-60.

Wei, Bifan, Jun Liu, Qinghua Zheng, Wei Zhang, Xiaoyu Fu and Boqin Fen. 2013. "A Survey of Faceted Search." *Journal of Web Engineering* 12: 41-64.

Willetts, Margaret. 1975. "An Investigation of the Nature of the Relation between Terms in Thesauri." *Journal of Documentation* 31: 158-84.

Wilson, Travis. 2006. "The Strict Faceted Classification Model." http://facetmap.com/pub/

Zelevinsky, V., J. Wang, and D. Tunkelang. 2008. "Supporting Exploratory Search for the ACM Digital Library." In *HCIR 2008: Proceedings of the Second Workshop on Human-Computer Interaction and Information Retrieval Microsoft Research, Redmond, Washington, USA October 23, 2008*, 85-88. http://ryenwhite.com/hcir2008/doc/HCIR08-Proceedings.pdf